AXES KIS/INS Interactive 2011

System Overview and Evaluation

Kevin McGuinness
Dublin City University

Robin Aly
University Twente

SEVENTH FRAMEWORK PROGRAMME

AXES

Overview

- System Overview
- User interface
- System design
- Experiments
- Future

System Overview

- Web browser-based user interface
- Search using:
  - Text
  - Images (visual similarity)
  - Concepts
- Text Search on Metadata and ASR
  - Apache Lucene 3.1.2
  - Five metadata fields: title, description, keywords, subject, uploader

System Overview

- Visual Concepts
  - 10 Concepts:
    - faces, female face, airplane, boat/ship, cityscape, singing, gender, nighttime, demonstration, playing instrument.
    - Subset of 5 used for INS
  - Pyramid histogram of visual words (PHOW) descriptor
    - Dense grid of VQ SIFT features at multiple resolutions
  - Ranked using non-linear χ^2 SVM
    - Trained using PEGASOS stochastic gradient descent algorithm (vlfeat implementation)
    - Train 100K frames in ~2 mins
    - Classify 100K frames in ~1 min

System Overview

- Visual Similarity Search
  - Web service that accepts a URL and returns a list of visually similar images
  - Based on "Video Google"
    - Hessian-affine interest points
    - SIFT descriptors quantized to visual words
    - Text retrieval methods on visual words
  - Search 100K frames in < 1 sec

System Overview

- Fusion of results
  - Simple weighted combination of results from text ASR search, text metadata search, visual concept search, and image similarity search
  - All scores (text, concepts, similarity) normalized to [0,1] by dividing through the max score
  - Active concepts equally weighted
  - The text, concept, and similarity scores equally weighted

$$score = \lambda_1 score_{text} + \frac{\lambda_2}{n} \sum_{i=1}^{n} score_{c_i} + \frac{\lambda_3}{m} \sum_{j=1}^{m} score_{sim_j}, \qquad (1)$$

User Interface

- Same user interface used for both KIS and INS tasks
- Web browser-based (Google Chrome only)
- Heavy emphasis on drag-and-drop
  - Drag to save shots
  - Drag to add shots to visual similarity search

TRECVID **KIS INTERACTIVE** | BY AXES TEAM

Find the video of people using ferry and touring ellis island

ellis island

FACES
BOAT OR SHIP
GENDER
PLAYING INSTRUMENT
FEMALE FACE
CITYSCAPE
NIGHTTIME
AIRPLANE
SINGING
DEMONSTRATION

GO ▶

EXAMPLE SHOTS

Drag shots to the left panel to use them in your query. Drag shots to the right panel to remove them from your query.

USE AS QUERY

QUERY

DON'T USE AS QUERY

SAVED SHOTS

Save as many relevant shots here as possible. Remove shots by dragging them to the "don't use as query" area or the trash can.

SAVED

1 2 3 4 5 6 7 8 9 10 11 12 13

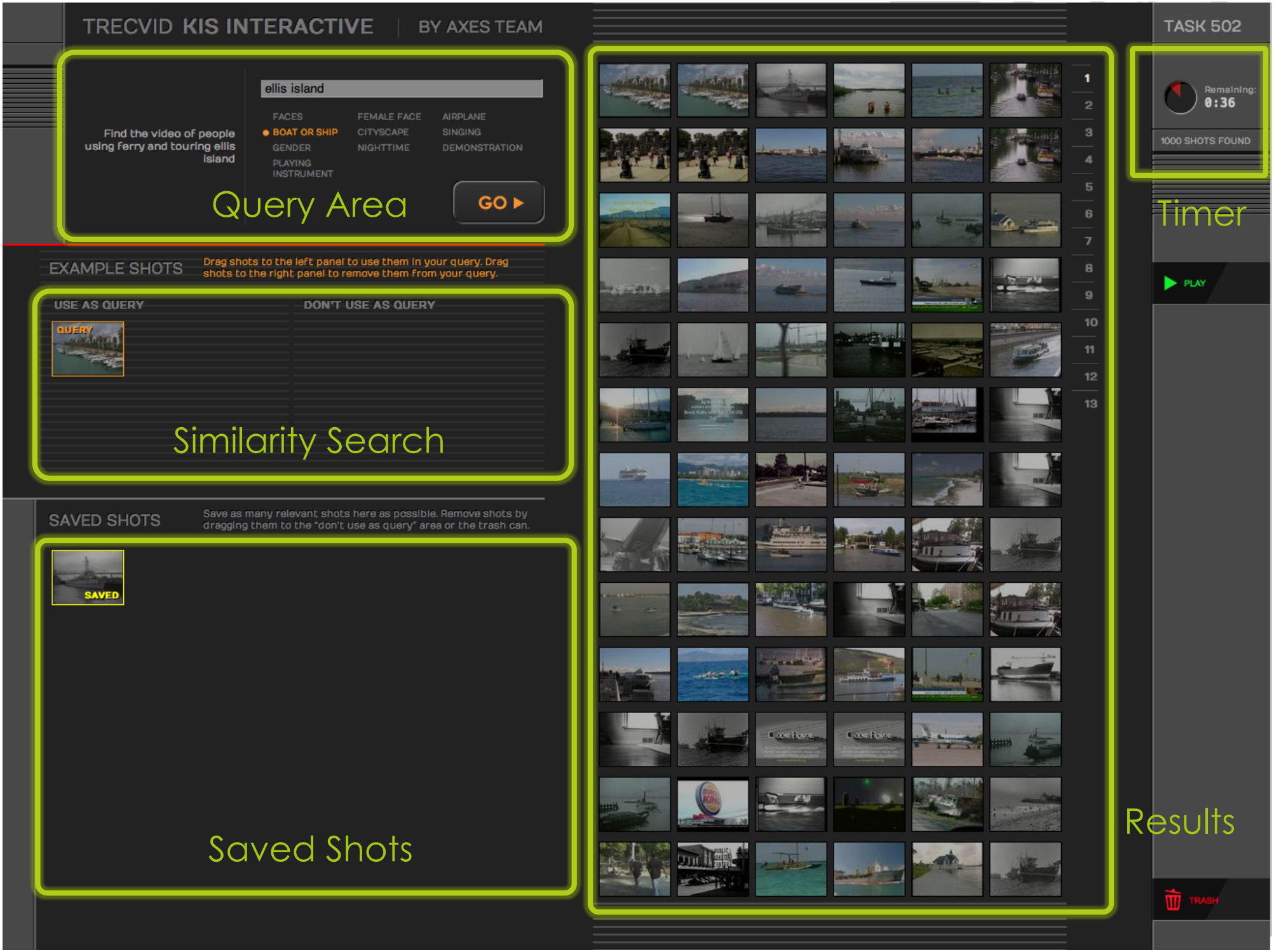
TASK 502

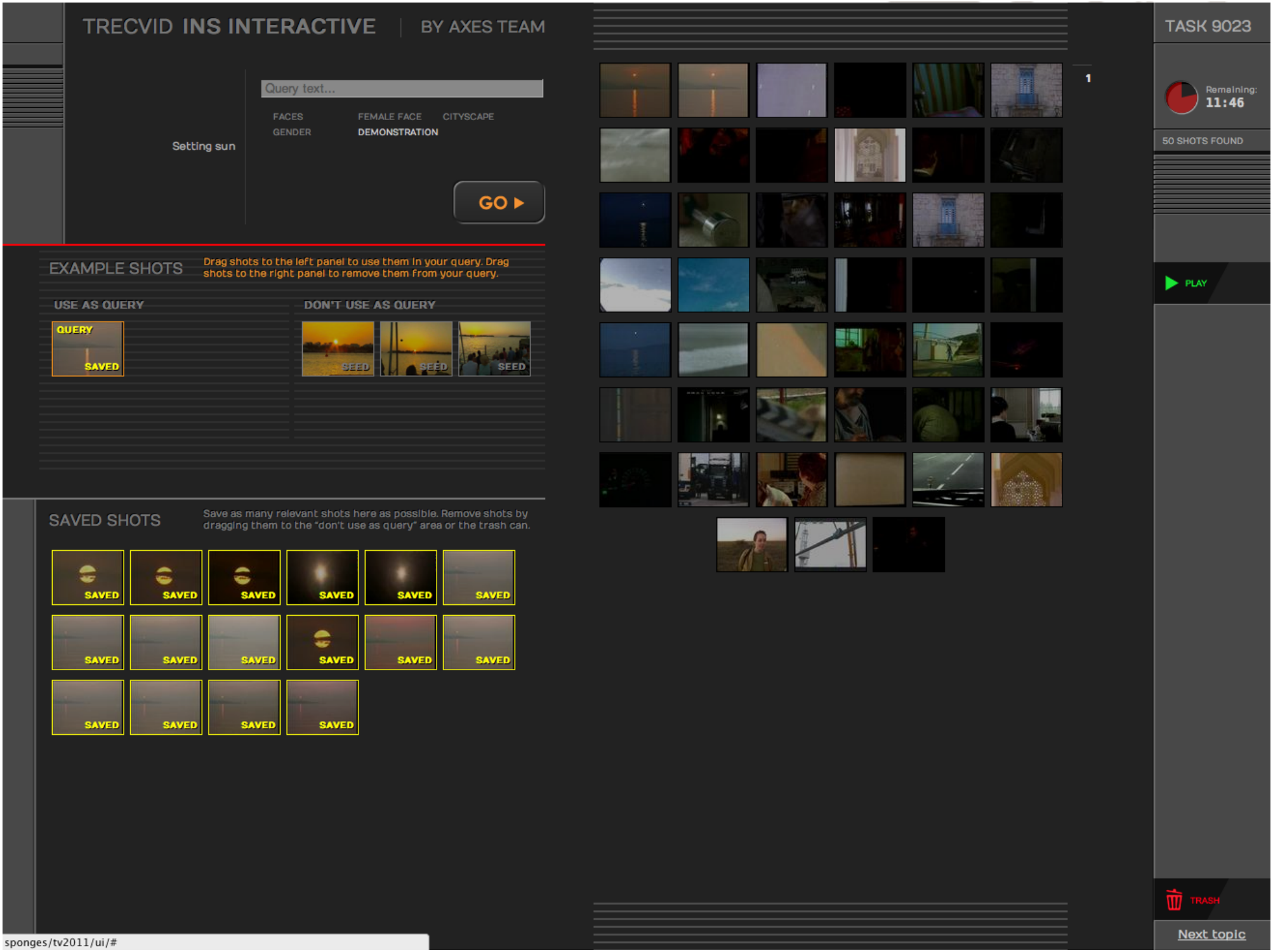
Remaining: 0:36

1000 SHOTS FOUND

PLAY

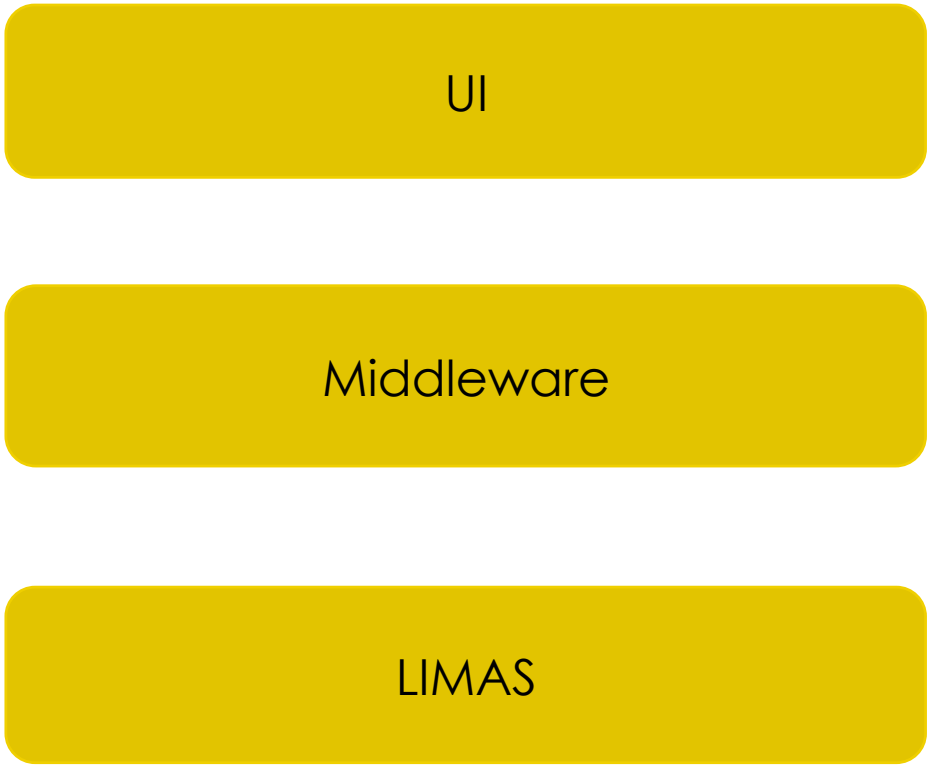
TRASH


TRECVID KIS INTERACTIVE
BY AXES TEAM
TASK 502
ellis island
Find the video of people using ferry and touring ellis island
FACES
BOAT OR SHIP
GENDER
PLAYING INSTRUMENT
FEMALE FACE
CITYSCAPE
NIGHTTIME
AIRPLANE
SINGING
DEMONSTRATION
GO
Query Area
Remaining:
0:36
1000 SHOTS FOUND
Timer
EXAMPLE SHOTS
Drag shots to the left panel to use them in your query. Drag shots to the right panel to remove them from your query.
USE AS QUERY
DON'T USE AS QUERY
QUERY
Similarity Search
PLAY
SAVED SHOTS
Save as many relevant shots here as possible. Remove shots by dragging them to the "don't use as query" area or the trash can.
SAVED
Saved Shots
1
2
3
4
5
6
7
8
9
10
11
12
13
Results
TRASH


TRECVID INS INTERACTIVE
BY AXES TEAM
Setting sun
Query text...
FACES
FEMALE FACE
CITYSCAPE
GENDER
DEMONSTRATION
GO
EXAMPLE SHOTS
Drag shots to the left panel to use them in your query. Drag shots to the right panel to remove them from your query.
USE AS QUERY
DON'T USE AS QUERY
QUERY
SAVED
SEED
SAVED SHOTS
Save as many relevant shots here as possible. Remove shots by dragging them to the "don't use as query" area or the trash can.
SAVED
1
TASK 9023
Remaining:
11:46
50 SHOTS FOUND
PLAY
TRASH
Next topic
sponges/tv2011/ui/#

Video Demo

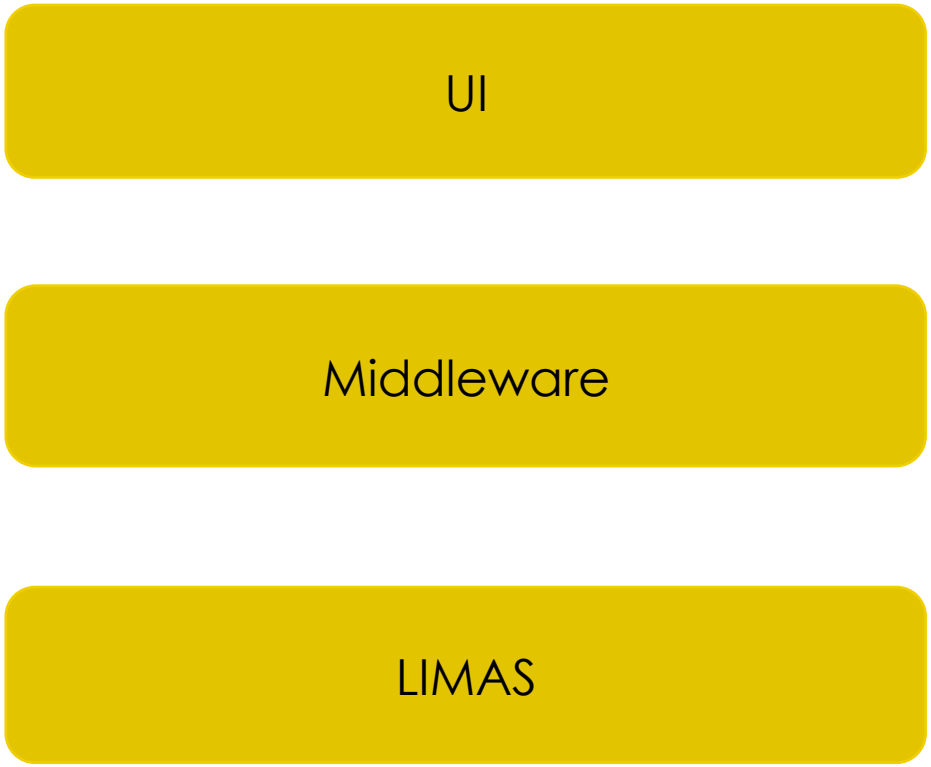
System Design



System Design

Responsibilities:

- Present tasks to user
- Allow user to formulate query
- Present results to user
- Time experiments
- Gather results



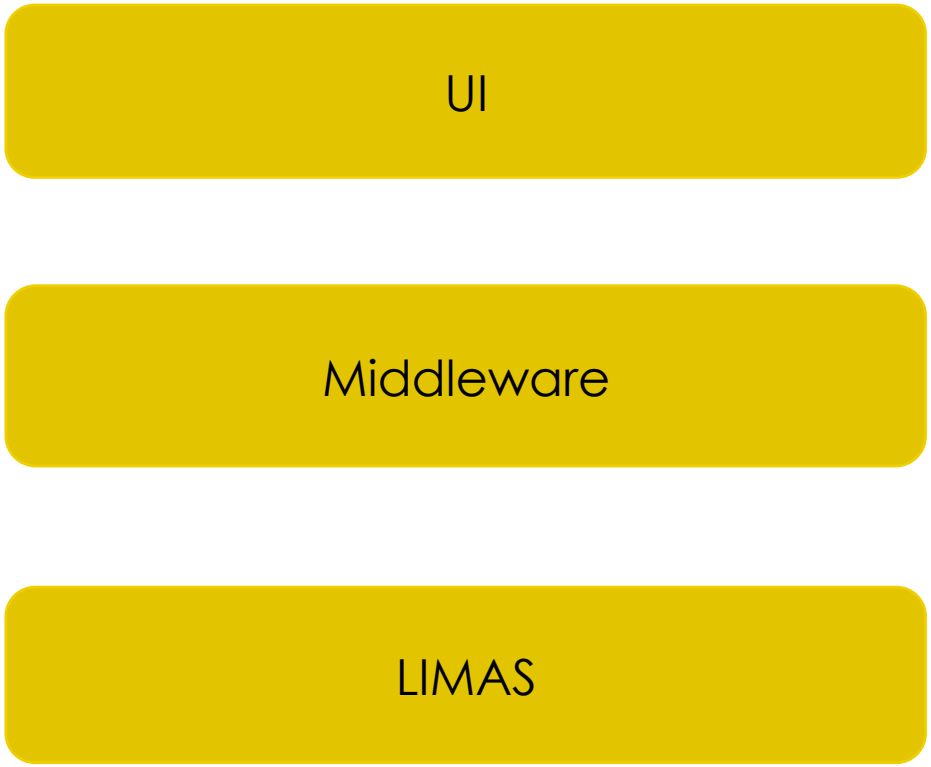
Technologies:

- HTML5
- CSS3
- Javascript
- JQuery
- AJAX

System Design

Responsibilities:

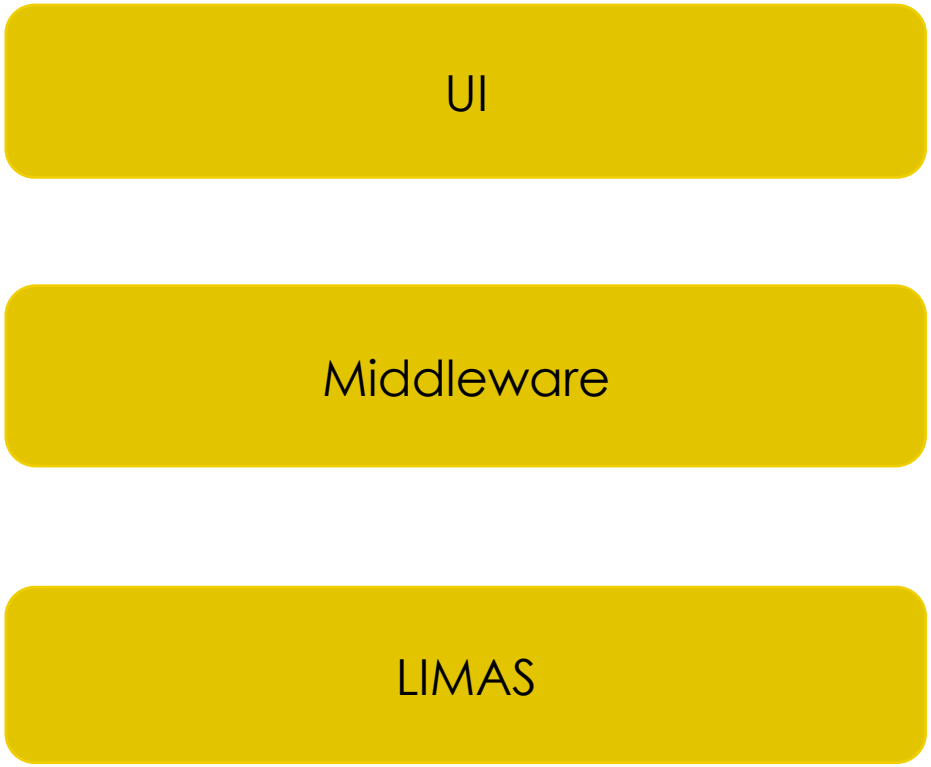
- Store topics, tasks, example images, etc. in a database
- Assign topics to users
- Mediate user queries
- Collect saved shots and store them in the database
- Log user actions
- Communicate with KIS oracle



Technologies:

- Python
- Django
- Apache/WSGI
- SQLite 3

System Design



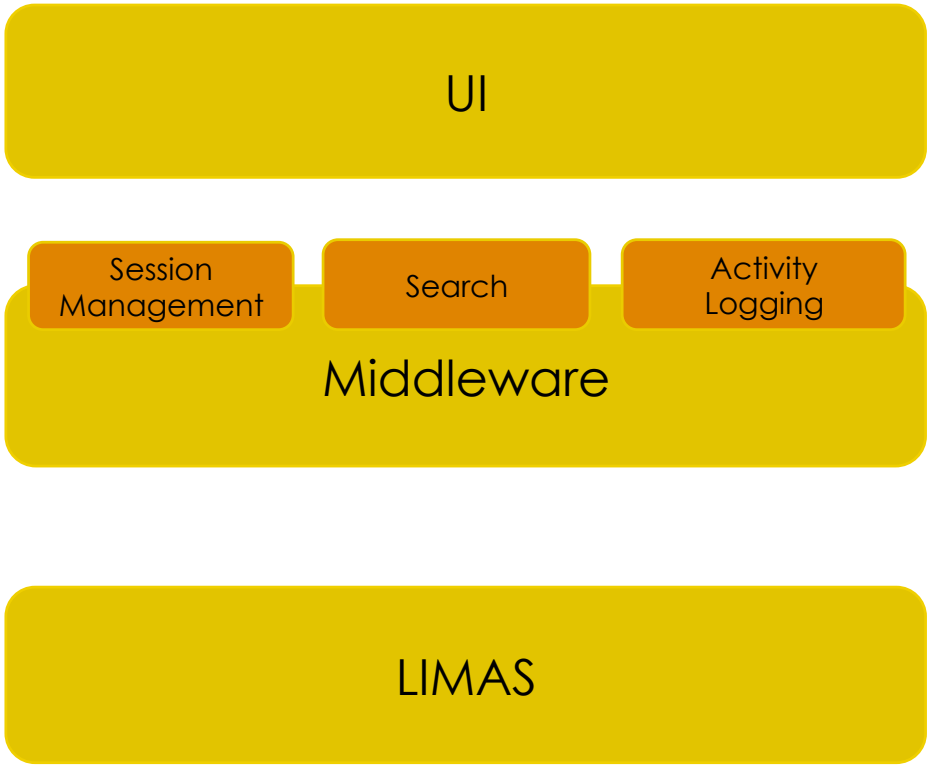
Responsibilities:

- Visual concept indexing and search
- Text indexing and search
- Communication with Oxford Similarity search
- Fusion of results

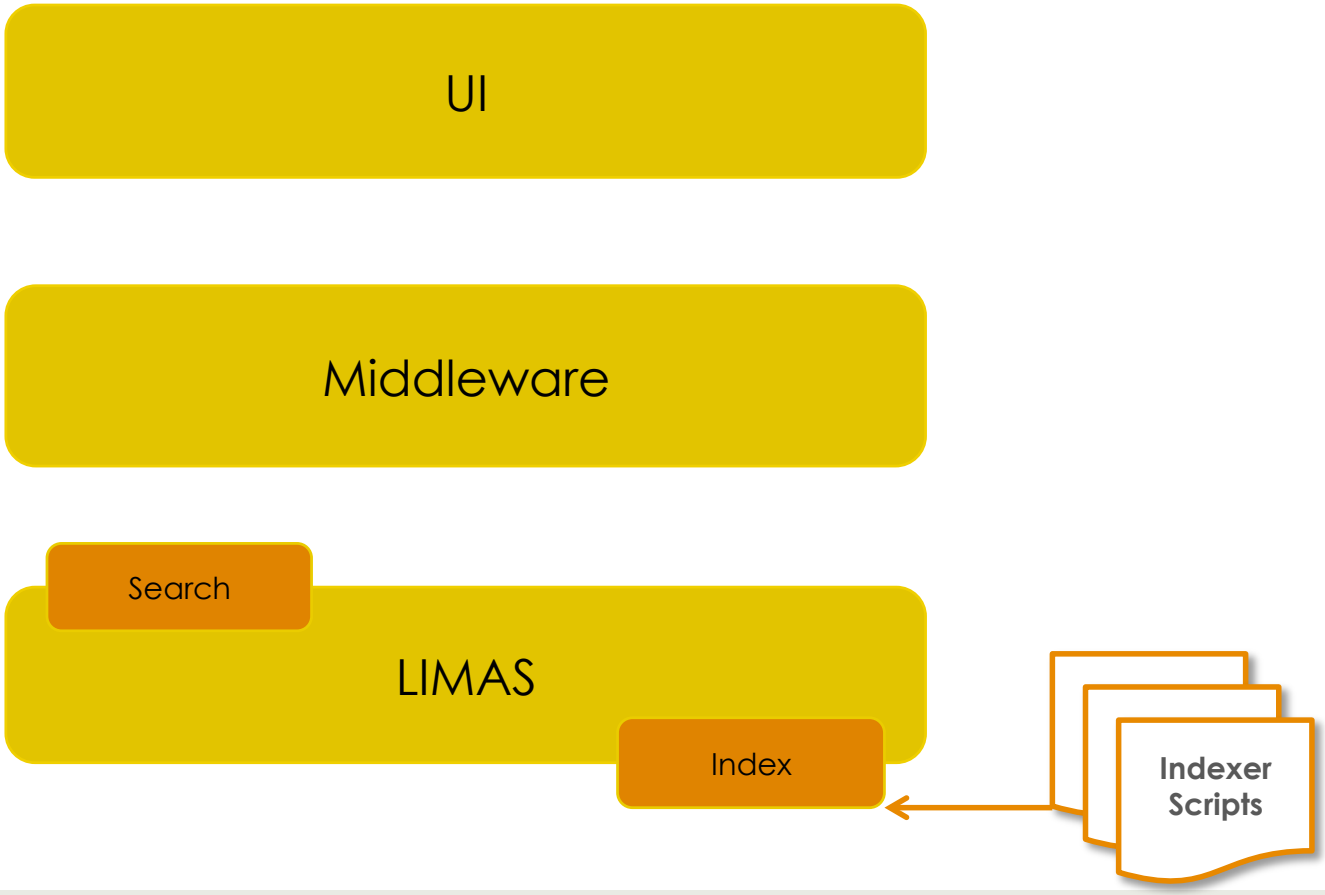
Technologies:

- Java
- Servlets
- Tomcat
- Apache Lucene
- Hadoop/HBase

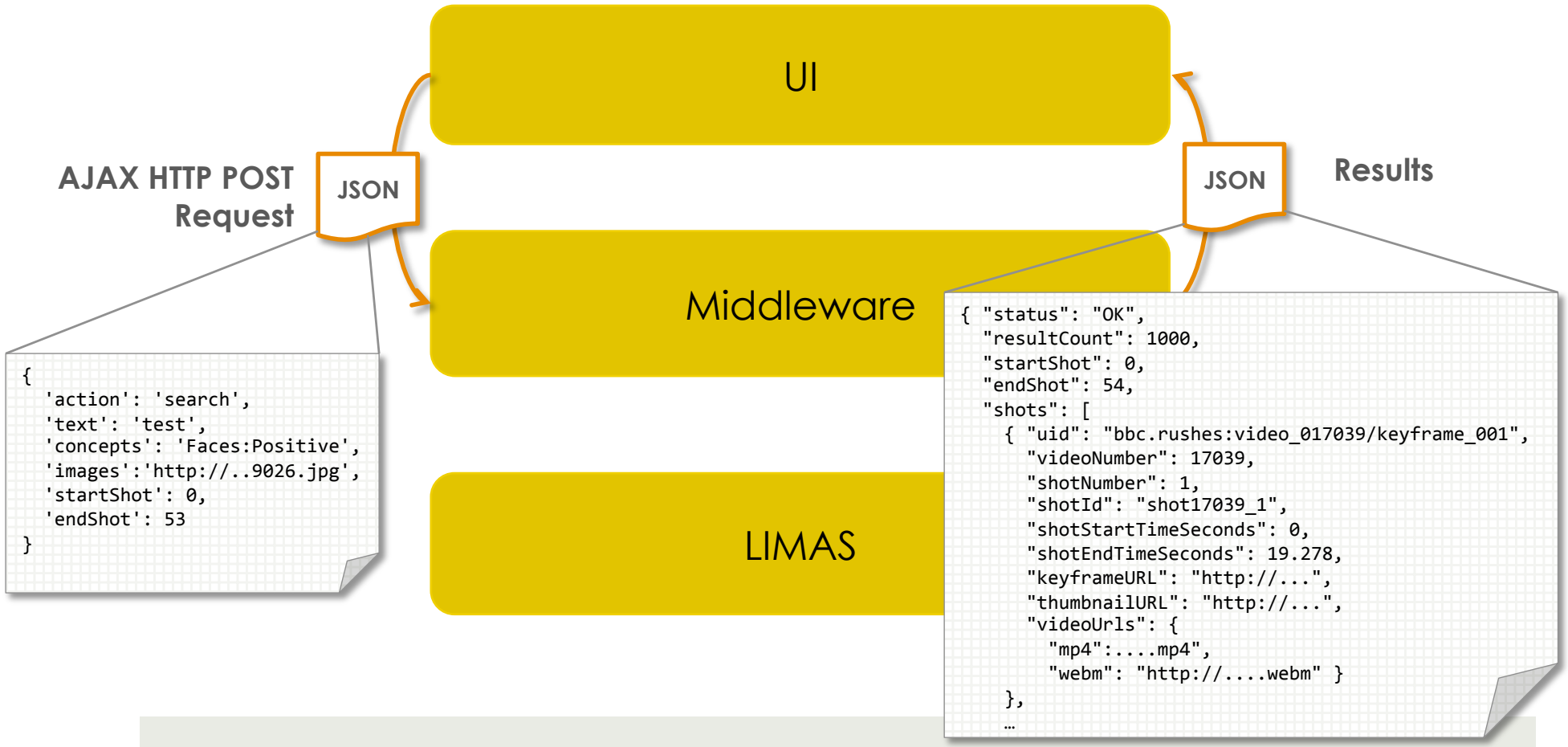
System Design



System Design

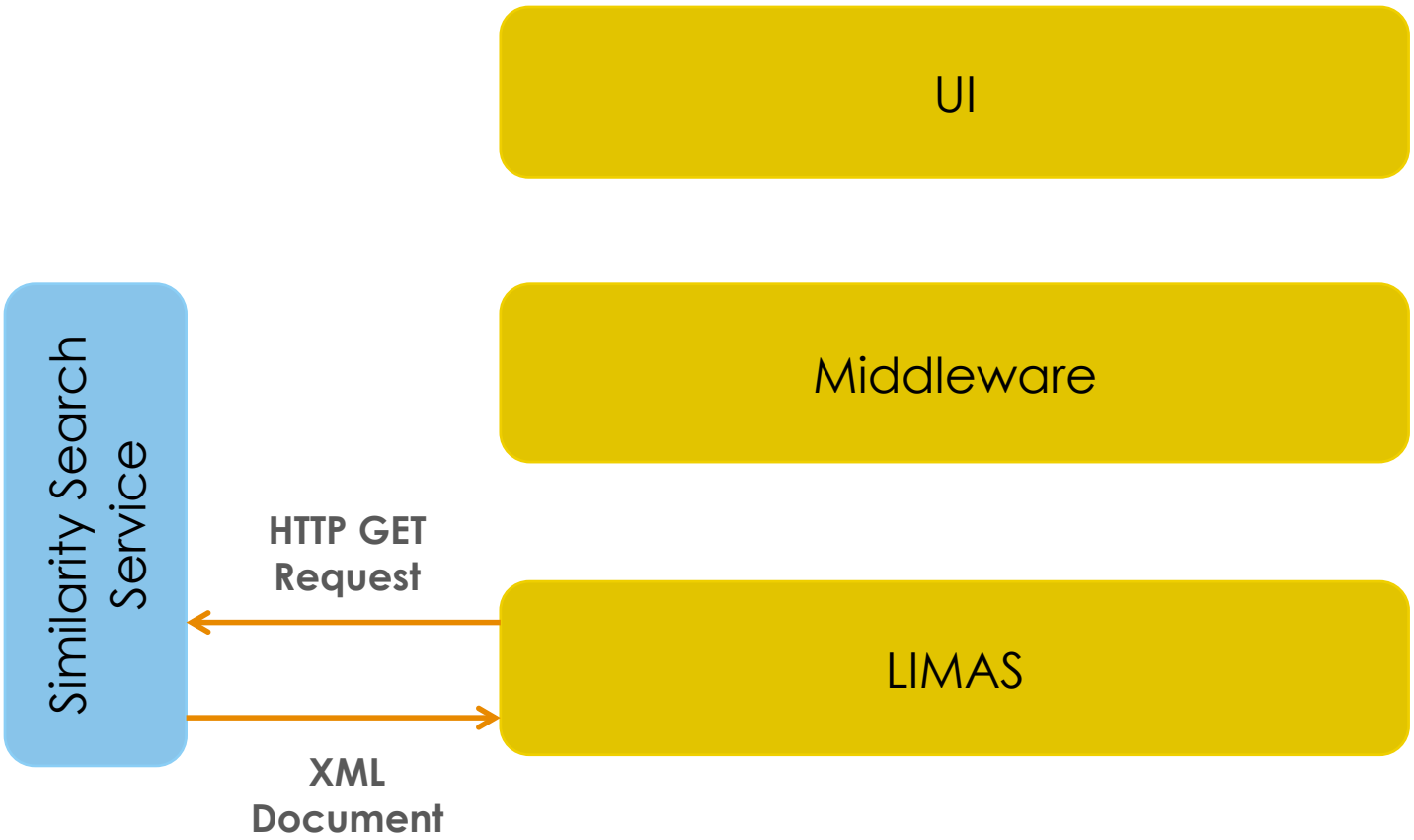


Communication



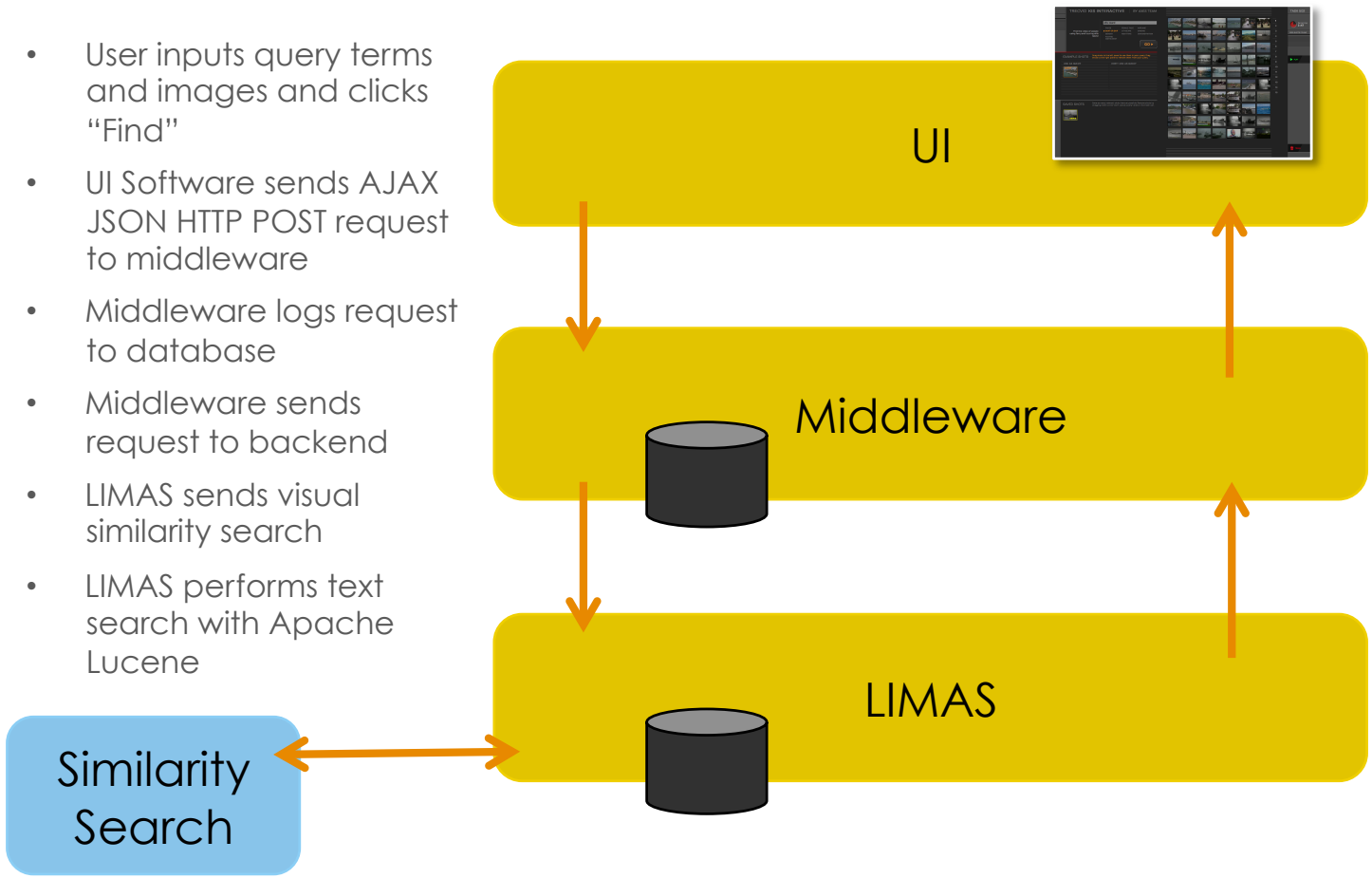
Communication

Communication



Typical Interaction

- User inputs query terms and images and clicks "Find"
- UI Software sends AJAX JSON HTTP POST request to middleware
- Middleware logs request to database
- Middleware sends request to backend
- LIMAS sends visual similarity search
- LIMAS performs text search with Apache Lucene



- LIMAS fuses results into a single result list
- LIMAS sends result list in JSON format to middleware
- Middleware logs results to database
- Middleware sends results in JSON format to UI
- UI Generates HTML for results and displays them to the user

Experiments

- NISV Hilversum, early September
- Known item search
  - 14 Media Professionals
  - 10 topics each
  - 5 minutes per topic (1 hr total)

- Instance search
  - 30 media students from Washington state (varying age)
  - 6 topics each
  - 15 minutes per topic (1.5 hr total)

Experiments

- Before experiment…
  - Participants briefed on purpose of experiment
  - Participants given short tutorial on UI
- After experiment…
  - Participants given freeform feedback form to fill out

The experiment setting

KIS Experiments

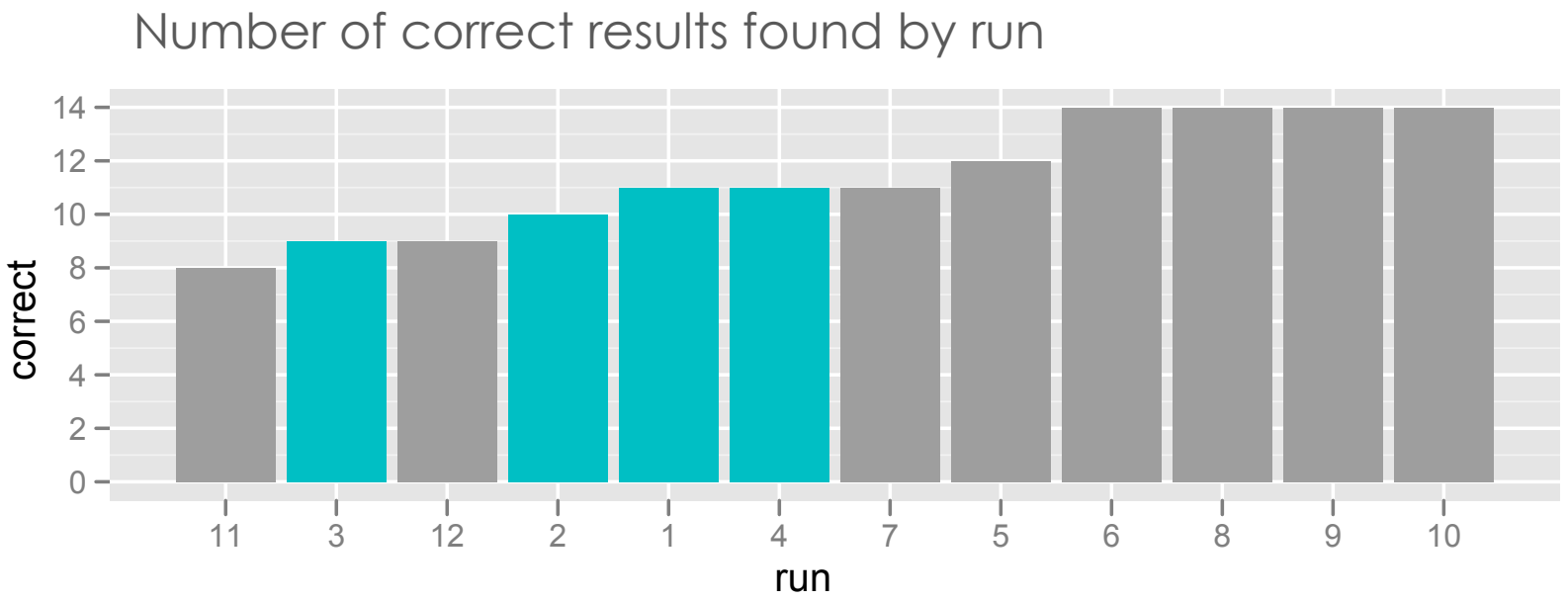
- 4 runs submitted
  - AXES_DCU_[1-4]
  - Same interface and system for all runs
  - Different users
  - Each user was randomly assigned to a single run

INS Experiments

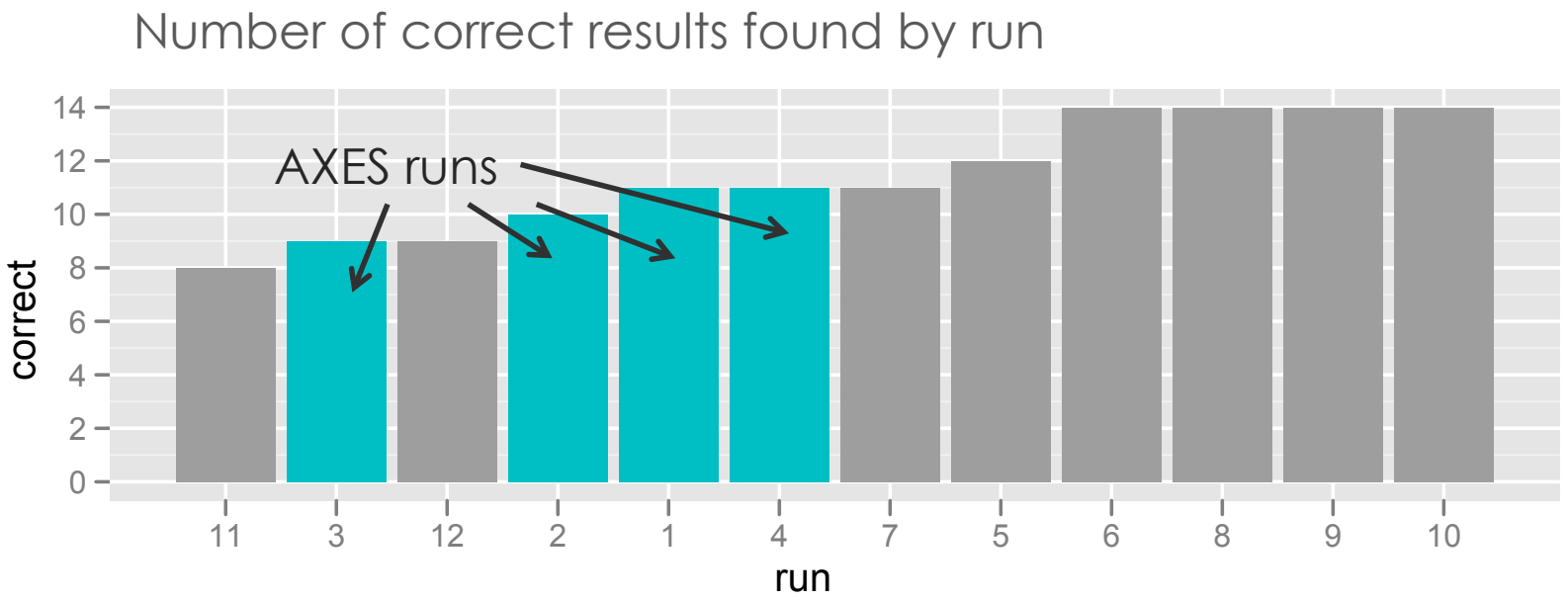
- 15 simultaneous users for INS experiments
  - Latin-square method
- Some technical issues during the experiments
- 4 runs ordered by the recall orientation of users
- Unfortunately, no other team participated

KIS Results

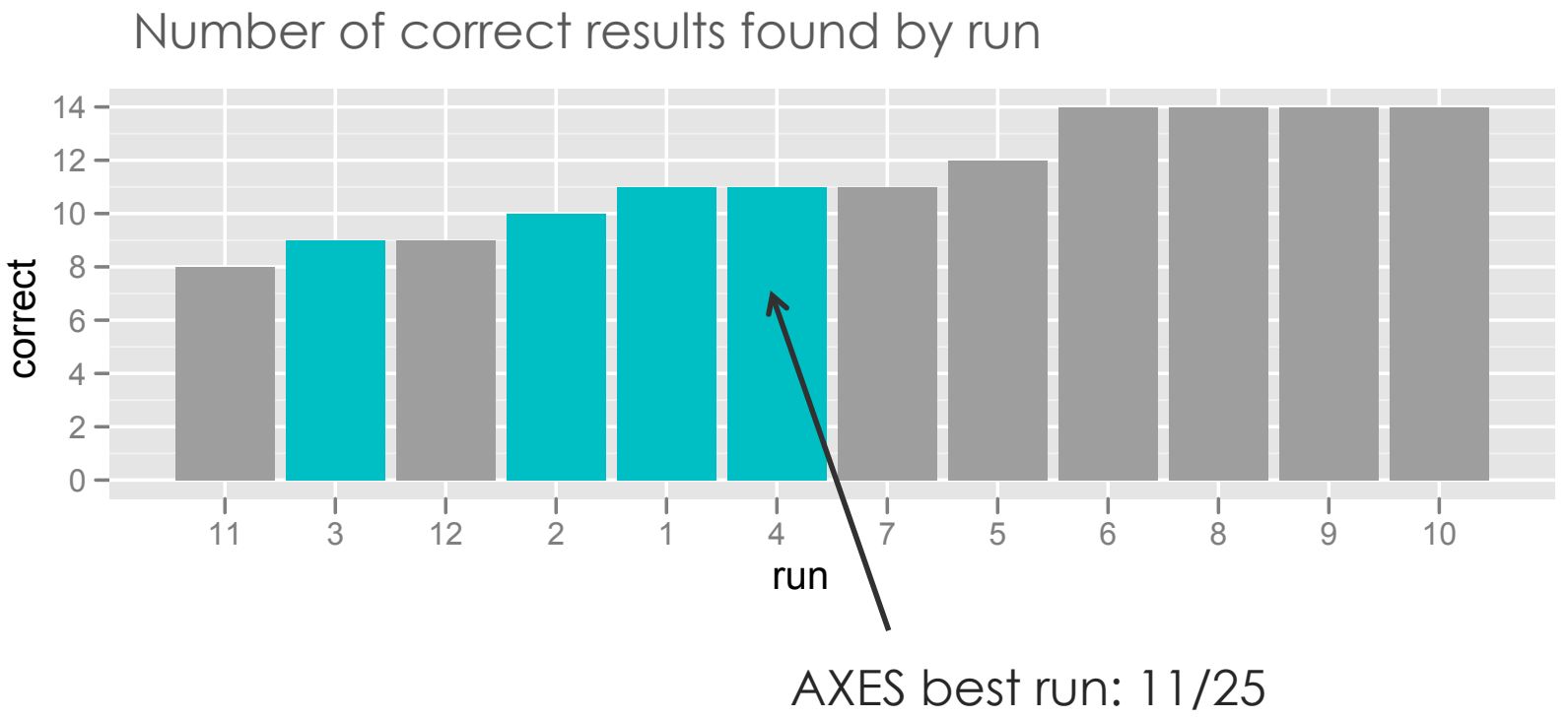
Evaluation (KIS)



Evaluation (KIS)



Evaluation (KIS)

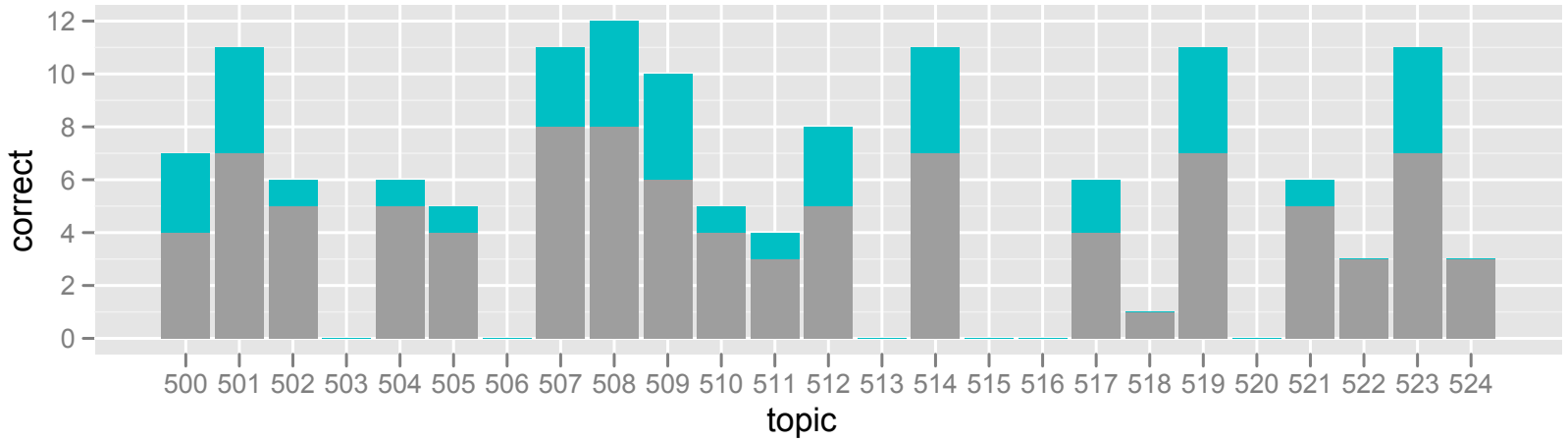


Evaluation (KIS)

AXES worst run: 9/25

Evaluation (KIS)

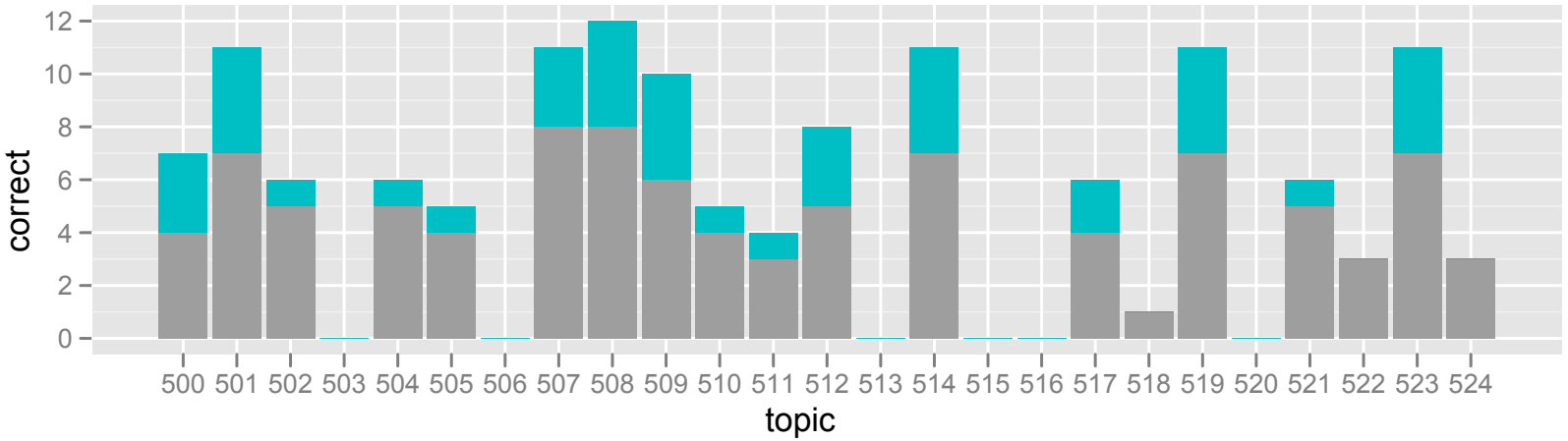
Number of correct results found by topic



Everybody found 501 and 508

Evaluation (KIS)

Number of correct results found by topic

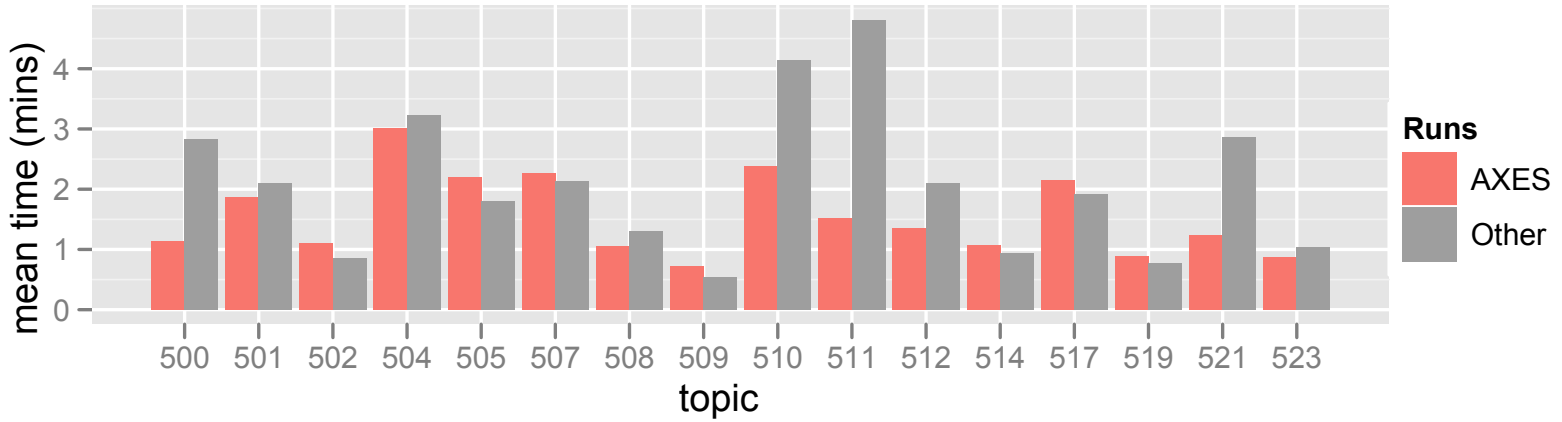


Everybody found 501 and 508

Nobody found 503, 505, 513, 515, 516, and 520

Evaluation (KIS)

Mean time to find the correct video by topic



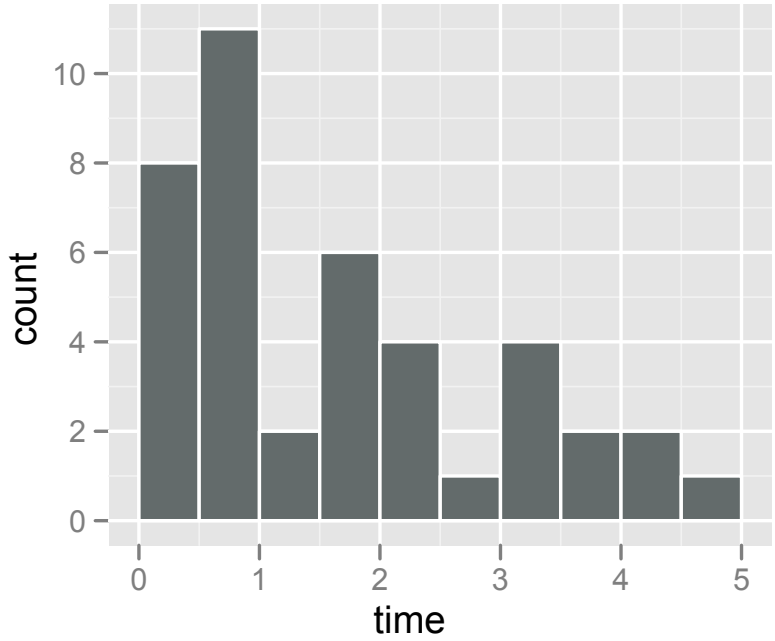
(Topics where the correct answer was not found by any AXES runs are not shown)

Evaluation (KIS)

Histogram of time taken to find the correct video (all runs)

19/41 (46%) of videos found were found in first minute

31/41 (75%) of videos found were found in first 2.5 minutes



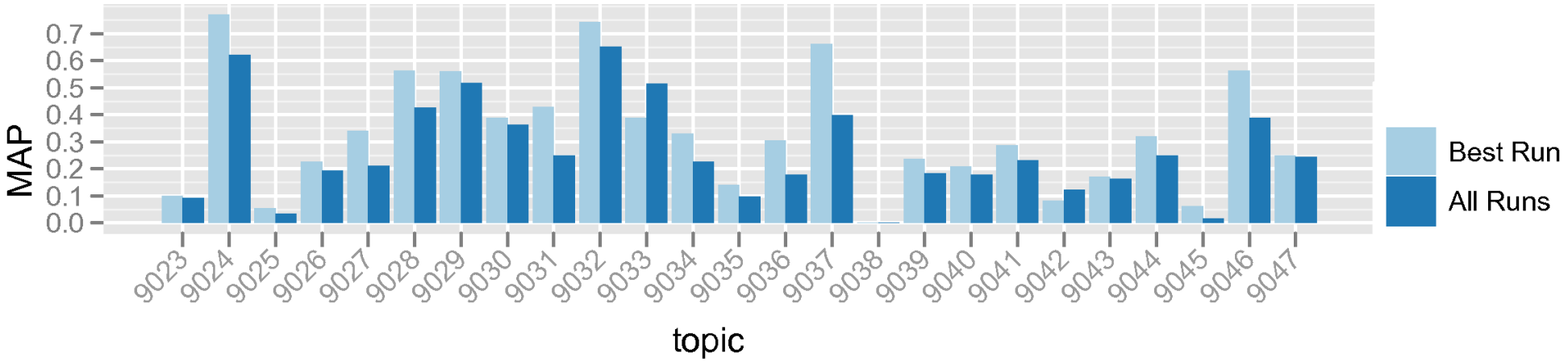
INS Results

Evaluation (INS)

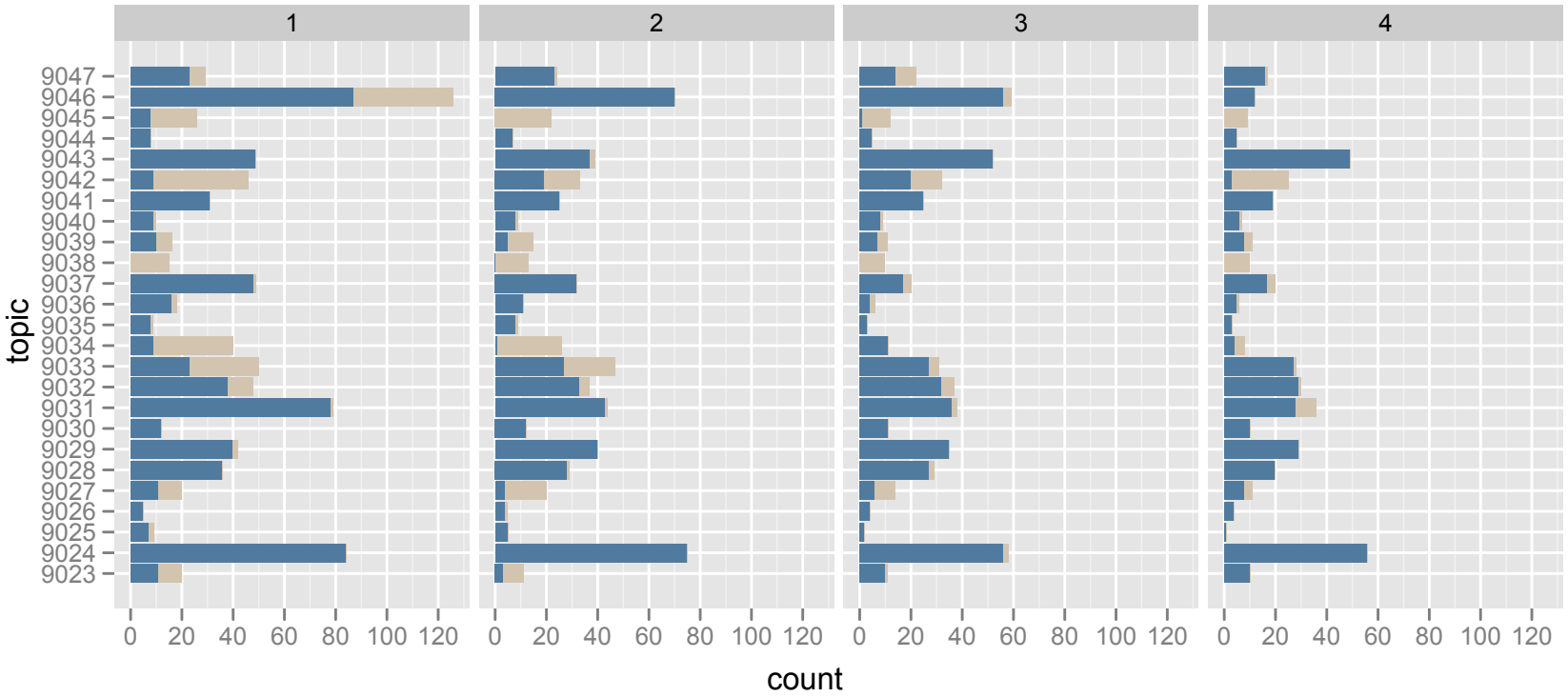
run	precision	recall	MAP	bpref	rel	non-rel
1	0.74	0.36	0.33	0.34	26.40	8.68
2	0.73	0.28	0.26	0.27	20.80	5.60
3	0.81	0.26	0.25	0.25	18.76	3.12
4	0.81	0.21	0.21	0.21	14.76	2.68

Evaluation (INS)

- Per topic comparison



Evaluation (INS)



Evaluation Summary

- Large variation in user performance!
  - For KIS a combined run containing our best performing users would have found 16/25 videos
  - Only 5/25 topics were found by all of our users
- Large variation in topic difficulty
  - Six topics found by no submitted run
  - Two topics found by all submitted runs
  - One topic only found by one submitted run
- Similar results from INS experiments

Feedback

- Users liked UI design and drag and drop based interaction mechanism
- Participants would have preferred to be able to adjust video size
- Professional users were unclear if Boolean search could be used
- Participants would like the system to give better hints on why a video was judged by the system to be relevant
  - Some remarked they did not know how the system worked and were not able to learn the system to adjust their search strategy

Feedback

- Users seemed to enjoy the task and the system ☺
- Lots of users said they wanted visual similarity search
  - Although, visual similarity was used less in the KIS task
- People used the visual concepts
- Got some great feedback from users
  - Excellent resource for building the future systems

Experiences

- Text is very important for KIS
  - If the metadata/ASR had some text that described the video, users usually found the correct one.
  - If there was no good metadata or ASR that matched the query topic, it's very hard to find the video using concepts and visual similarity alone

Conclusions

- Participation of AXES in the KIS & INS Task
- Simple Fusion Approach of Similarity, Concepts and ASR
- Known-item search
  - 14 media professionals participated
  - Median performance (MAP)
- Instance search
  - 30 media students from Washington participated
  - Only task participant in INS
- Users were positive about possibilities

Future

- TRECVid 2012
  - Improve fusion
  - UI enhancements based on user feedback
  - Pre-clustering results on video

Questions?